A Reconfigurable 16/32 Gb/s Dual-Mode NRZ/PAM4 SerDes in 65-nm CMOS

Ashkan Roshan-Zamir, Student Member, IEEE, Osama Elhaddity, Member, IEEE, Hae-Woong Yang, Student Member, IEEE, and Samuel Palermo, Member, IEEE

Abstract—While four-level pulse amplitude modulation (PAM4) standards are emerging to increase bandwidth density, the majority of standards use simple binary non-return-to-zero (NRZ) signaling. This paper presents a dual-mode NRZ/PAM4 serial I/O SerDes which can support both modulations with minimum power and hardware overhead relative to a dedicated PAM4 link. A source-series-terminated transmitter achieves 1.2-Vpp output swing and employs lookup table control of a 31-segment output digital-to-analog converter (DAC) to implement 4/2-tap feed-forward equalization in NRZ/PAM4 modes, respectively. Transmitter power is improved with low-overhead analog impedance control in the DAC cells and a quarter-rate serializer based on a tri-state inverter-based mux with dynamic pre-driver gates. The receiver implements an NRZ/PAM4 decision feedback equalizer that employs one finite impulse response and two infinite impulse response taps for first post-cursor and long-tail inter-symbol interference (ISI) cancellation, respectively. First post-cursor ISI cancellation is performed in these comparators to optimize the design’s timing, while the remaining ISI taps are subtracted in a preceding current integration summer for improved sensitivity. Fabricated in GP 65-nm CMOS, the transceiver occupies 0.074 mm² area and achieves 16 Gb/s NRZ and 32 Gb/s PAM4 operation at 10.9 and 5.5 mW/Gb/s, respectively. 

Index Terms—Decision feedback equalizer (DFE), dual-mode serial link, feed-forward equalizer (FFE), impedance tuning, infinite impulse response (IIR), non-return-to-zero (NRZ), pulse amplitude modulation (PAM4), receiver, transmitter.

I. INTRODUCTION

IMPROVEMENTS in high-speed serial I/O bandwidth density and energy efficiency are necessary to support the dramatic growth in global IP traffic, which is projected to reach 2 zettabytes per year by 2019 [1]. While high-performance I/O circuitry can leverage technology improvements, unfortunately the bandwidth of the electrical channels used for inter-chip communication has not scaled in the same manner. This merits serious consideration of four-level pulse amplitude modulation (PAM4) which, relative to simple binary non-return-to-zero (NRZ) signaling, offers higher spectral efficiency, lower loss at the Nyquist frequency, and relaxed clock speeds. These advantages have led to the implementation of PAM4 modulation in various high-speed I/O standards [2], [3]. In order to support PAM4 modulation, there have been recent developments in current-mode [4]–[7], voltage-mode [8], and hybrid transmitters [9], and both analog-to-digital converter (ADC) based [7], [10], [11] and mixed-signal receivers [5], [6], [12]. Relative to NRZ-based systems, PAM4 transceivers require more stringent circuit linearity, equalizers which can implement multi-level inter-symbol interference (ISI) cancellation, and improved sensitivity.

On the transmitter side, source-series-terminated (SST) voltage-mode drivers enable the high output swing required for PAM4 modulation with high linearity achieved up to differential output swings equal to the nominal output stage supply [13]. Further improvements in output swing are possible with advanced hybrid drivers employing current boosting [9]. Voltage mode drivers also offer reduced static power consumption relative to current-mode drivers. Although, at higher data rates this static power advantage becomes a smaller percentage of the total transmitter power consumption. Key reasons for this include large clocking power and that these voltage-mode drivers often use output-stage segmentation to achieve equalization setting and impedance control. The presence of equalization tap-select muxes that must pass the full-rate signal in the output segments [13] can introduce on-chip ISI and including digitally controlled redundant segments for impedance control [14] results in increased output stage area and power. Another key transmitter bottleneck is the final serializer, where efforts have been made to minimize power consumption in both current-mode [15] and voltage-mode [16] implementations.

Equalization is often also implemented at the receiver to support higher channel loss, with the most common blocks employed being a continuous-time linear equalizer (CTLE) and a decision feedback equalizer (DFE). CTLE is effective at cancelling both pre-cursor and long-tail ISI. However, CTLE amplifiers must be designed with sufficient bandwidth to support the full rate signal and linearity to support PAM4 modulation. DFE is often used due to the effectiveness of cancelling ISI without amplifying noise or crosstalk [17]. However, a key challenge associated with DFE architectures involves optimizing the critical feedback path to allow for ISI cancellation.
beginning at the first post-cursor. While PAM4 modulation allows for a longer unit interval (UI) time, the reduced voltage margins necessitate increased comparator gain to achieve a symbol decision in one UI. Another issue is that DFEs which employ common finite impulse response (FIR) feedback filters can require a large tap count (> 10) to cancel long-tail ISI [18]. An efficient solution for this is to employ infinite impulse response (IIR) feedback filters which can cancel smooth exponentially decaying ISI with a minimal number of taps [17], [19]–[21], in a manner similar to a continuous-time equalizer. Finally, a PAM4 DFE must implement the necessary hardware with the required linearity to support multi-level ISI subtraction.

While PAM4 has better spectral efficiency relative to NRZ signaling, this does not make it the superior modulation option for all systems. The optimal modulation is a function of the target data rate, channel loss profile, and process technology, with the majority of standards utilizing simple binary NRZ signaling. As serial I/O transceivers are often designed to support different channels and standards, this motivates dual-mode transceivers with flexible equalization (Fig. 1) to seamlessly support both NRZ and PAM4 modulation with minimal hardware and power overhead.

This paper presents a quarter-rate 16/32 Gb/s dual-mode NRZ/PAM4 SerDes datapath which can be configured to work in both modes with minimal hardware overhead relative to a dedicated PAM4 link [22]. Section II investigates the equalization requirements of the proposed transceiver with statistical bit error rate (BER) modeling results of transmit-side feed-forward equalization (FFE) used with receive-side DFE structures with either FIR or IIR feedback taps. The high-swing voltage-mode SST transmitter which utilizes an efficient tri-state inverter-based mux with dynamic pre-driver gates, a lookup table (LUT) controlled 31-segment output digital-to-analog converter (DAC) to implement FFE without any full-rate tap-select muxes, and low-overhead analog impedance control is detailed in Section III. Section IV discusses the receiver that saves power with a quarter-rate DFE that directly samples the input from the termination and achieves efficient equalization with 1-FIR tap for the large first post-cursor ISI and 2-IIR taps for long-tail ISI cancellation [19]. Experimental results from a general purpose (GP) 65-nm CMOS prototype are presented in Section V. Finally, Section VI concludes the paper.

II. TRANSCEIVER ARCHITECTURE

High-speed link signal integrity suffers from ISI caused by channel skin effect, dielectric loss, and reflections. The proposed transceiver is designed to support refined electrical channels with minimal performance degradation due to reflections, such as the one shown in Fig. 2(a) which displays a smooth low-pass frequency response and 13.5 dB loss at 8 GHz. This causes attenuation and dispersion of a 16 GS/s data pulse at the channel output. The resultant time-domain ISI in Fig. 2(b) is well characterized by a fast rising side with only one significant pre-cursor ISI term, a fast-decaying short-tail ISI term that dominates through the third post-cursor, and a slow-decaying long-tail ISI term that continues out to higher post cursor locations [17], [20], [21]. While the first pre-cursor ISI term is small, it can significantly degrade performance in PAM4 systems due to this modulation being more sensitive to residual ISI, as further quantified in the Appendix. Thus, transmitter FFE should be utilized to cancel this pre-cursor term and RX DFE can compensate for the post-cursor terms.

The slow-decaying long-tail ISI can have a large impact and necessitate a large tap count in DFEs with conventional FIR feedback filters [18]. Utilizing the 16 GS/s pulse response in a statistical BER simulator, the 32 Gb/s PAM4 timing margin is compared in Fig. 2(c) assuming a 2-tap TX FFE for pre-cursor cancellation and various configurations of RX DFE feedback filters. While four FIR DFE taps can achieve a BER $< 10^{-12}$, nine FIR DFE taps are required to achieve an eye opening close to 10% at this BER. DFEs with IIR taps have been shown to efficiently cancel smooth exponentially decaying ISI, with only one IIR-tap utilized for signaling over an RC-limited on-chip channel [23]. However, a major issue with DFE IIR feedback taps is that the comparator regeneration can limit the time available for the IIR filter output to reach the required amplitude to cancel the large first-post cursor ISI term. This motivates hybrid DFE architectures which employ one FIR feedback tap for the first post-cursor ISI and subsequent IIR taps for long-tail ISI cancellation [21], [24]. Fig. 2(c) shows that by employing one FIR and one IIR feedback tap, a performance better than five FIR taps is achieved with 32 Gb/s PAM4 modulation. Multiple IIR feedback taps provide more flexibility to tailor the tap time constants and post-cursor location to better match a given printed circuit board (PCB) channel [17], [25], with close to 10% eye opening at BER $= 10^{-12}$ achieved by employing one FIR and two IIR taps.
Fig. 2. Refined electrical channel. (a) $S_{21}$ response. (b) 16GS/s pulse response. (c) 32 Gb/s PAM4 timing margin with 2-tap pre-cursor TX FFE and various RX DFE feedback filter configurations. (d) 16GS/s pulse response with 2-tap TX FFE and RX DFE with 1-FIR and 2-IIR feedback taps.

The pulses responses of Fig. 2(d) also confirm that this equalization configuration is effective in cancelling both pre-cursor and post-cursor ISI.

Fig. 3 shows the proposed dual-mode NRZ/PAM4 transceiver architecture. At the transmitter side, a modulation mode signal selects either a 1/16th or 1/8th symbol-rate clock to
control the 16-bit wide PRBS15 pattern generator and initial serialization stages in NRZ and PAM4 mode, respectively, to generate four sets of four-bit patterns which address the LUT equalizer that controls the 31-segment high-swing SST output stage. This allows the realization of a 4/2-tap FFE in NRZ/PAM4 mode, respectively. Relative to a dedicated PAM4 transmitter, the overhead to support NRZ modulation with the same 16-bit parallel interface consists of the additional 8:4 serializer stage, some additional latches in the modulation selector block, and the initial clock mux. Overall, these blocks are 4.2% of the total area and consume 3.6% of the total transmitter power. At the receiver side, a quarter-rate 3-tap NRZ/PAM4 DFE is utilized with 1 FIR and 2 IIR feedback taps. The three output bits per quarter-rate slice, which are all the same value for NRZ and thermometer-code for PAM4, are converted to binary and buffered out of the chip for BER testing. While not implemented in this prototype, a calibration scheme can be utilized similar to [26] to automatically correct phase mismatches and provide uniform output eyes.

A. 4-to-1 Serializer

The final 4-to-1 serializer is one of the most critical blocks in a quarter-rate transmitter, as it must maintain enough bandwidth to support the full-rate output. However, this can be difficult to achieve with conventional pass-gate serializers which suffer from reduced drive strength due to the effective transistor stacking at the high self-loading output node. This transmitter extends the 2-to-1 tri-state inverter-based mux design proposed in [27] to perform 4-to-1 serialization and further improves power efficiency by utilizing dynamic NAND pre-drivers [Fig. 5(a)]. Fig. 5(b) shows the serializer’s PMOS-path timing diagram, with similar waveforms present in the NMOS path. The dynamic NAND predriver gates utilize the input data to qualify a pulse defined by adjacent quarter-rate clock edges. This allows the tri-state inverter-based mux to drive the full-rate output node through only a single transistor, similar to a simple inverter, with the input data activating one of the PMOS/NMOS devices. Dummy gates are present in both the PMOS and NMOS paths to enable a uniform eye diagram at the full-rate serializer output. As shown in the
post-layout simulation results of Fig. 5(c), the proposed dynamic tri-state inverter-based design has significantly faster transition times relative to a conventional pass-gate serializer designed with equal power consumption. Overall, the minimal transistor stacking allows the proposed serializer to achieve the same level of deterministic jitter with a 40% power reduction relative to a conventional pass-gate design.

B. Pseudo-Analog-Controlled Output Driver

Fig. 6(a) shows a single segment of a conventional high-swing SST output driver. The segment’s output impedance is set by the series combination of the passive resistor, $R_{\text{term}}$, and the transistor’s triode resistance. As shown in the simulation results of Fig. 6(c), both of the elements are affected by process variations and can cause deviations in the driver output impedance without any compensation (no comp). A straightforward technique to control the output impedance of a high-swing voltage-mode driver involves implementing redundant segments that can be digitally activated to match the channel impedance [14]. However, the presence of these redundant stages results in increased output stage area and pre-driver power.

This design proposes pseudo-analog control to compensate for large statistical variations in driver output impedance. Fig. 6(b) shows a schematic of the voltage-mode SST driver segments which supports a 1.2-V$_{\text{pp}}$ output swing. Here the main MP and MN switch transistors and $R_{\text{term}}$ resistors are sized to always yield greater than 50-$\Omega$ output impedance over corners, and two analog-controlled paths are added for impedance tuning via the GP/N gate voltages. While conceivably one additional analog-controlled branch is sufficient for impedance control, a tradeoff exists in choosing $R_P$ and $R_N$ values. As shown in Fig. 6(c), selecting a relatively small $R_P$ and $R_N$ value to yield near 50-$\Omega$ under a $+3\sigma$ variation case (Single leg comp1) results in low overdrive voltages for the MRP and MRN transistors under a nominal impedance corner. This causes a large positive deviation from the desired 50-$\Omega$ value due to the transistors entering the saturation
region with a small-signal output impedance higher than the large-signal value set by a conventional analog control loop. Conversely, selecting a relatively high RP and RN value to yield near 50 \( \Omega \) under a nominal variation case (Single leg comp2) results in insufficient overdrive voltage range and a large positive deviation under a +3\( \sigma \) impedance corner. Thus, in order to break this tradeoff, a two branch compensation approach with both analog-controlled low-impedance path 1 and high-impedance path 2 are added which are replica-biased by the finite-state machine (FSM)-controlled pseudo-analog loop.

Fig. 7 shows the output driver impedance controller that produces the output voltages, GP1, GN1, GP2, and GN2, that control the low/high-impedance paths’ pull-up and pull-down resistances. The impedance controller consists of a replica transmitter stage with a precision off-chip 100-\( \Omega \) resistor load that is placed in two feedback loops. Depending on the control-loop mode, the top loop sets the MRP1/2 transistors’ gate voltage with either the analog control signal VOP or in a digital fashion to be fully-on (VSS) or fully-off (VDD) in order to force a value of \((3/4) \times VDD\) at the replica transmitter positive output. The bottom loop works in a similar manner to force a value of \((1/4) \times VDD\) at the replica transmitter negative output. For corners with low output resistance, the impedance tuning circuitry operates with the lower-impedance path 1 in the feedback loops to set analog voltages GP1/GN1 with VOP/VON to yield a 50 \( \Omega \) match, while the higher-impedance path 2 is disabled (Mode 1). In Mode 1, both the replica driver and the main output driver segments share the same control signals. For corners with high output resistance, path 1 switches from analog to digital control and is turned fully on, while the higher-impedance path 2 is now in the feedback loops to set analog voltages GP2/GN2 with VOP/VON to yield a 50 \( \Omega \) match (Mode 3). In Mode 3, again both the replica driver and the main output driver segments share the same control signals. For corners with close-to-nominal output resistance, the main output driver is designed to operate with path 1 simply set fully on and path 2 disabled, while the replica loop controls either the low- or high-impedance path depending on the previous state (Mode 2A/B). Switching between the modes in the replica loop without dithering the control signals presented to the main output driver is achieved by an asynchronous FSM that monitors the VON voltage. As shown in the Fig. 7 flowchart, in the nominal impedance case the replica driver will be continuously switching between the low-impedance (Mode 2A) and high-impedance (Mode 2B) modes without disturbing the output driver segments. In Mode 2A with the low-impedance path in feedback, the loop checks whether VON is less than a high threshold \( VH \), corresponding to deep triode operation of MRN1, minus some margin before transitioning to Mode 1 with analog control of the low-impedance path 1 in the main output stage segments. In Mode 2B with the low-impedance path fully on and the high-impedance path in feedback, the loop checks whether
VON is greater than a low threshold VL, corresponding to a minimum conductance level from MRN2, plus some margin before transitioning to Mode 3 with analog control of the high-impedance path 2 and the low-impedance path 1 fully on in the main output stage segments. The margin introduced in transitioning between Modes 1/2 and 2/3 introduces hysteresis which, along with the extra Mode 2 state, prevents dithering in the main output segment impedance control signals.

Robust operation of the replica bias impedance control scheme in the presence of mismatch is ensured since the analog-controlled MRP and MRN transistors are always biased in the triode region with a large overdrive voltage. In order to quantify the effect of both process variation and mismatch between the replica and output driver segments, the output driver’s post-layout simulated return loss plot is shown in Fig. 8 with ±3σ error bars. While there is some slight variation over the process corners, the small error bars indicate that the mismatch-induced variation for a given corner is minimal. Overall, the simulation results show a worst case return loss of −27.4 and −10.5 dB at 500 MHz and 8 GHz, respectively.

IV. RECEIVER

Fig. 9 shows the proposed dual-mode NRZ/PAM4 DFE receiver block diagram. In PAM4 mode symbol detection is achieved with a 2-bit flash ADC consisting of three comparators with threshold voltages of 0, ±2/3 relative to the post-equalized differential amplitude, while in NRZ mode all thresholds are set to zero. A quarter-rate architecture is employed to reduce clock buffer power and allow for longer comparator reset time, which minimizes hysteresis and allows smaller pre-charge transistor loading for improved evaluation delay. In order to minimize the critical first-tap feedback delay and maximize the equalization cancellation range, an FIR tap is utilized to cancel the first post-cursor ISI. This multi-level FIR tap is efficiently realized by feeding back the flash ADC 3-bit thermometer-coded output bits directly to three equally-weighted summer inputs embedded in the comparators’ first stage, removing any SR-latch and external summer delay from this critical path. Long-tail ISI is efficiently cancelled with 2 IIR taps, with one tap starting from the second post-cursor to cancel fast time constant ISI and the other beginning at the third post-cursor to mitigate the slow time constant ISI. In order to minimize the comparator’s internal loading, these IIR taps are subtracted from the sampled input with a current integration summer that precedes the comparators.

A detailed timing diagram of the NRZ/PAM4 DFE receiver is shown in Fig. 10, which includes voltage waveforms for ideal PAM4 data and highlights a decision made by the comparator bank clocked by phase 0. First an input sample-and-hold tracks the input signal for 2UIs starting at the rising edge of clock phase 90 and holds the sampled value for 2UIs when clock phase 270 rises. For 1UI between the rising edges of clocks 270 and 0, the current integrating summer preceding the 2-bit flash ADC 0 comparator bank sums the input signal with the outputs from both the first IIR1 filter, which is currently driven by the 180 comparator bank outputs, and the second IIR2 filter, which is currently driven by the 90 comparator bank outputs. During this time, the 270 comparator bank is regenerating to make the previous symbol decision. A symbol decision is made on the rising edge of phase 0 with the final first post-cursor DFE FIR tap realized via integration in the 0 bank comparators with the 270 bank comparators thermometer outputs subtracting from the current integrating summer output. This quarter-rate architecture provides an additional benefit of maintaining the comparator output levels for an extra UI relative to a half-rate design, which relaxes the comparator hold time requirements and allows direct connection to the IIR1 filter between the rising edges of clocks 90 and 180. At the output of the comparator bank, an SR latch preforms RZ to NRZ conversion to produce the final thermometer outputs which also drive the IIR2 filter between the rising edges of clocks 180 and 270.

A. Comparator and DFE FIR Tap

The comparator is one of the most critical building blocks in DFE receivers because it determines the maximum data-rate and overall sensitivity, which sets the maximum loss that the DFE can compensate. While common strong-arm comparators [28], and modified double-tail versions [23] [Fig. 11(a)], have advantages that include no dc power, small aperture time, high gain, and CMOS-level outputs, they generally have larger delays relative to dynamic latches [29]–[31]. However, as shown in Fig. 11(b), these dynamic latches typically have relatively low gain because of the absence of a regeneration stage, which impacts their robustness with reduced PAM4 voltage margins.

A simplified (latch-only) schematic of a newly proposed regenerative latch consisting of a two-stage dynamic amplifier, similar to a dynamic latch, but with a regeneration stage connected in parallel with the second stage is shown in Fig. 11(c). This additional parallel stage increases the latch gain relative to a dynamic latch and generates CMOS-level outputs. Cross-coupled inverters are utilized, versus only cross-coupled PMOS transistors, for robust operation with variations in the first-stage common-mode output. The clocked NMOS tail transistor prevents static power consumption in the pre-charge phase and regulates the discharge current through the transistors Mn3 that start at the beginning of
the regeneration phase. Relative to a double-tail latch which requires complementary clocks to activate the second-stage PMOS tail transistor, utilizing an NMOS tail transistor in the proposed latch allows for operation with only a single clock phase and reduced loading on the clock buffers. In order to reduce the second-stage loading and minimize the effect of the mismatches in the cross-coupled inverters on the overall comparator offset, the cross-coupled inverters are designed to be half the size of the second-stage main amplifier.

Fig. 12 shows how the comparators first stage is augmented to include an additional differential stage for offset correction/threshold control and three differential inputs controlled by the flash ADC 3-bit thermometer outputs for FIR-tap summation. In order to minimize loading on the first-stage output node, the current from these FIR-tap pairs is connected through the Mn4 transistor whose gate voltage $V_{FIRb}$ is controlled by a 7-bit DAC to set the tap weighting. This Mn4 transistor acts as a source-degenerated current source whose current is controlled by the $V_{FIRb}$ voltage, which is decoupled well for noise rejection. While the Mn5–7 differential pair transistors should be sized large enough for low overdrive voltages and fast current switching, the statically-controlled Mn4 transistors can be sized smaller to yield a 15% reduction in the total output drain capacitance relative to direct connection of the Mn5–7 transistors. The FIR tap range is 180 mV with the 3-bit thermometer decision feedback signals hardwired to implement a negative first post-cursor tap, which is generally desired in wireline applications.
Fig. 11. Comparator comparisons. (a) Double-tail latch. (b) Dynamic latch. (c) Proposed latch.

Fig. 12. (a) Modified first stage of the proposed latch to include PAM4 FIR-tap summation and offset correction/threshold control. Comparator input offset versus (b) $V_{FIR_b}$ and (c) $V_{OFF}/V_{th}$. (d) Normalized summation weight versus input amplitude at 16 GS/s.

A manually-tuned 7-bit DAC also controls the offset correction/threshold control differential input $V_{OFF}/V_{th}$ to allow for over a ±400 mV range. Post-layout simulations show that an input amplitude of 20 mV is required at 16 GS/s to achieve 90% FIR correction and avoid noise propagation in the DFE loop [32].

The comparator performance is summarized in the additional post-layout simulation results of Fig. 13. Utilizing the nominal 1-V supply and a 0.7-V common-mode input level, the comparators’ delay is near 75 and 51 ps for input amplitudes of 5 and 20 mV, respectively, to allow for either 12.5 or 16 GS/s operation. Assuming a 35-mV input amplitude, the delay varies less than 10% if the common mode is maintained above 0.7 V. The impulse sensitivity function of the comparator shows an aperture time of 12 ps with maximum sensitivity at 8 ps after the clock’s rising edge. Given that the comparators power is mostly dynamic, the cumulative power efficiency for all the comparators in the receiver is relatively
constant over clock frequency at near 0.16 mW/Gbps. The comparator occupies an area near 100 $\mu$m$^2$ and has an input-referred offset sigma of 12 mV and noise of 0.6 mV$_{\text{rms}}$.

B. Current Integrating Summer

Preceding the 2-bit flash ADC comparator banks are the current integrating summers shown in the top portion of Fig. 14, which are used to subtract the IIR taps from the input signal. Input sample-and-holds (S/Hs) are employed in the DFE slices to mitigate frequency dependent loss [33]. These are implemented with PMOS transistors only because of the high common mode level and include dummy transistors to cancel charge injection and input feed-through during the hold time [34]. A simulated 1 dB compression point of 480-mV input amplitude provides sufficient linearity to support PAM4 modulation. The S/H output feeds the summers main input whose gain/linearity is controlled through varying the split-tail current sources and the degeneration resistance between 50–300 $\mu$A and 500 $\Omega$–3.5 k$\Omega$, respectively, providing gain tuning from 0 to 6 dB. This allows the receiver to operate at different data rates with a wide range of input amplitudes. The other summer inputs are connected to the two IIR taps’ outputs. As their maximum amplitude is relatively small, no degeneration resistors are used for these inputs. The clocked transistors Mn4 and Mp1 are used to allow current integration over only a single UI, which prevents the IIR filters output corresponding to the following bits from affecting the summer output. PMOS current injectors are utilized at the summer output to provide common-mode restoration for proper operation of the comparators in the subsequent 2-bit flash ADC [18], [34]. The summer integrates for 1-UI only and holds its output for the following UI while the comparator is in the sampling and regeneration phases, which improves the sensitivity.

C. DFE IIR Taps

As shown in the bottom portion of Fig. 14, in order to implement the IIR taps the quarter-rate flash ADC 3-bit outputs are multiplexed to full-rate using three identical parallel current-mode multiplexers whose outputs sum onto an RC filter. The filter time constant is controlled though varying the resistance $R_d$ and capacitance $C_d$, while the output amplitude is set by adjusting $I_{\text{tap}}$. This ensures that the output common mode is fixed and only a function of $R_0$ and $I_0$. The first IIR filter time constant can be adjusted by changing $C_d$ and $R_d$ with 3-bit and 2-bit digital-control, respectively. This results in an IIR1 time constant tuning range between 0.5 UI and 4.5 UI at 16 GS/s, as shown in Fig. 15(a). The second IIR filter utilizes the same 2-bit digitally-controlled $R_d$, but expands the $C_d$ control to 5-bits to realize longer time constants. This results in an IIR2 time constant tuning range between 0.5 and 10 UI at 16 GS/s, as shown in Fig. 15(b). For both IIR taps, the tail current is implemented with 5-bit digital control.

D. Sensitivity Analysis

As discussed in Section I, receiver sensitivity is important to support the reduced level separation and increased impact of residual ISI with PAM4 modulation. Key receiver sensitivity components include the front-end circuitry input-referred noise and the minimum input amplitude necessary for the comparators to make a sufficient decision. Table I summarizes the simulated noise contributions of the main DFE blocks, with the comparator noise dominating the 0.73 mV$_{\text{rms}}$ total input-referred noise. Considering this total noise and the minimum 20 mV differential input amplitude necessary for 90%
summer correction, the DFE has sensitivity of 50.3 mV_{ppd} per PAM4 eye for a BER of $10^{-12}$ at 16 GS/s.

![Image of circuit diagram]

**Fig. 14.** Current integrating summer and DFE IIR1 tap details.

**Fig. 15.** Time constant of the (a) first IIR filter and (b) second IIR filter versus the tuning code.

**TABLE I**

<table>
<thead>
<tr>
<th>Block</th>
<th>Noise Contribution (mV_{ppd})</th>
</tr>
</thead>
<tbody>
<tr>
<td>S/H</td>
<td>0.35</td>
</tr>
<tr>
<td>CI Summer</td>
<td>0.24</td>
</tr>
<tr>
<td>Comparator</td>
<td>0.6</td>
</tr>
<tr>
<td>Total</td>
<td>0.73</td>
</tr>
</tbody>
</table>

V. EXPERIMENTAL RESULTS

The dual-mode NRZ/PAM4 SerDes was fabricated in a 65-nm CMOS GP process. As shown in the die micrographs of Fig. 16, the total active area for the transmitter is 0.06 mm$^2$ and the DFE receiver core is 0.014 mm$^2$. The four-phase clocks for the quarter-rate SerDes are generated on both chips by passing a half-rate differential input clock through on-chip CML divide-by-2 blocks followed by CML-to-CMOS converters and local clock buffers.

Fig. 17 shows measurement results of the transmitter positive and negative output pin impedance versus output differen-
Fig. 16. Chip micrograph of (a) transmitter and (b) receiver.

Fig. 17. Measured transmitter output impedance versus differential output voltage for (a) positive output pin and (b) negative output pin.

Fig. 18. Level separation mismatch ratio (RLM) measurement results for (a) nominal PAM4 level settings and (b) optimized PAM4 level settings.

tial voltage for five different transmitter chips. The impedance control loop ensures that the output stage maintains near a 50-Ω output impedance over the entire 1.2 Vpp range for both nominal samples (1–4) which operate in Mode 2 and the high-impedance variation sample 5 which operates in the analog-controlled Mode 3. Fig. 18 shows level separation mismatch ratio (RLM) measurements which highlight the utility of the LUT-based transmitter. Utilizing the default PAM4 settings result in a 93% RLM, with the third level being somewhat low in this sample. Optimizing the LUT settings allows for an improved 96.7% RLM and more uniform level spacing.

A block diagram of the link BER test setup and measurements of the two test channels’ insertion loss is shown are Fig. 19. Eye diagrams are captured at the output of the test channels, excluding the RX PCB loss of about 3 dB at 8 GHz, utilizing a high-bandwidth sampling scope to characterize the transmitter. Full link testing is performed with two synchronized sources to generate the transmitter and receiver clocks. A programmable-phase shifter is inserted in the receive-side path to manually adjust the phase and generate BER bathtub curves. This receive-side clock is also used to clock the BERT. In PAM4 mode, the on-die quarter-rate data MUX at the receiver output allows for independent verification of the MSB or LSB outputs. These results are then combined to produce the receiver BER bathtub curves in PAM4 mode.
The transmitter eye diagrams at the channels’ outputs and the full-link BER timing margin bathtub curves are shown in Figs. 20 and 21, respectively. Fig. 21 also includes the utilized TX and RX equalizer settings, with initial values
obtained using the statistical simulation model discussed in Section II and further manual fine tuning employed to achieve the lowest BER. 32 Gb/s PAM4 operation is achieved over channel 1, which has 13.5 dB loss at the 8 GHz Nyquist frequency. The left half of Fig. 20 shows that without any transmit equalization the output eye diagram is completely closed. As shown in Fig. 21(a), utilizing only RX equalization in this case allows for only a BER near $10^{-10}$. While only optimizing the PAM4 2-tap TX FFE allows for open eyes at the channel output before the RX PCB, the additional board loss results in only 0.02 UI timing margin at a BER $= 10^{-12}$ without any receiver equalization. Co-optimizing the 2-tap TX FFE for pre-cursor ISI cancellation with the RX DFE for post-cursor cancellation allows this timing margin to increase to 0.06 UI. Note that in this co-optimized condition the eye diagram at the RX PCB input is completely closed, as shown in Fig. 20(c). 16 Gb/s NRZ operation is achieved over channel 2, which has 27.6 dB loss at the 8 GHz Nyquist frequency. The right half of Fig. 20 shows that without any transmit equalization the output eye diagram is completely closed. As shown in Fig. 21(b), utilizing only RX equalization in this case allows for only a BER near $10^{-8}$. Optimizing the NRZ 4-tap TX FFE allows for open eyes at the channel output before the RX PCB, as depicted in Fig. 20(e). Jitter decomposition of this eye yields 34.3 ps of deterministic jitter, with 31.2 ps of residual ISI being the main contributor. The random jitter is measured at 830 fs rms using a clock source with 750 fs rms random jitter. A timing margin of 0.08 UI at a BER $= 10^{-12}$ is achieved without any receiver equalization. This timing margin is improved to 0.18UI with co-optimization of the 4-tap DX FFE and RX DFE. As in the PAM4 case, in this co-optimized NRZ condition the eye diagram at the RX PCB input is completely closed, as shown in Fig. 20(f).
TABLE II
TRANSEIVER PERFORMANCE SUMMARY

<table>
<thead>
<tr>
<th>References</th>
<th>Data Rate (Gb/s)</th>
<th>Equalization</th>
<th>Modulation</th>
<th>Total Loss @ Nyquist (dB)</th>
<th>Eye Width BER</th>
</tr>
</thead>
<tbody>
<tr>
<td>This Work</td>
<td>32/16</td>
<td>2-tap TX FFE + 1-tap FIR, 2-tap IIR RX DFE</td>
<td>PAM4</td>
<td>13.5</td>
<td>6% 10^{-12}</td>
</tr>
<tr>
<td>[5]</td>
<td>20</td>
<td>4-tap TX FFE + 1-tap FIR, 2-tap IIR RX DFE</td>
<td>PAM4</td>
<td>27.6</td>
<td>18% 10^{-12}</td>
</tr>
<tr>
<td>[6]</td>
<td>56</td>
<td>3-tap TX FFE</td>
<td>PAM4</td>
<td>5</td>
<td>-</td>
</tr>
<tr>
<td>[35]</td>
<td>36</td>
<td>3-tap TX FFE + RX CTLE + ADC based RX 24-tap FFE, 1-tap DFE</td>
<td>NRZ</td>
<td>25dB</td>
<td>-</td>
</tr>
<tr>
<td>[34]</td>
<td>28</td>
<td>5-tap TX FFE + RX CTLE + 14-tap RX DFE</td>
<td>NRZ</td>
<td>40 dB for 25.78 Gb/s</td>
<td>23% 10^{-12}</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>3-tap TX FFE + RX CTLE + 14-tap RX DFE</td>
<td>NRZ</td>
<td>34 dB</td>
<td>- 10^{-15}</td>
</tr>
</tbody>
</table>

*Clock generation and CDR power included

Fig. 22 shows the 32 Gb/s power breakdown. The transmitter consumes 158.6 mW of power, with the 4-to-1 serializer and local clock buffers having the most contribution. Only 17.7 mW is consumed at the receiver, with the local clock buffers and comparators dominating. Table II summarizes the multi-mode transceiver performance and compares this work against other dedicated NRZ and PAM4 designs. Relative to the mixed-signal PAM4 designs of [5] and [6], the presented transceiver’s additional equalization functionality allow for compensation of higher channel loss. Better power efficiency and significant area reduction is also achieved relative to a 16-nm ADC-based PAM4 transceiver [35]. Comparing against NRZ designs, the presented dual-mode transceiver achieves a higher 32 Gb/s data rate in PAM4 mode at a better power efficiency than a 28-nm design operating at 28 Gb/s NRZ [13]. Also, superior power efficiency in NRZ operation is achieved relative to the 16 Gb/s 40-nm design which utilizes a DFE with 14 FIR feedback taps [36].

VI. CONCLUSION

This paper has presented a 16/32 Gb/s dual-mode NRZ/PAM4 SerDes which can be configured to work in both modes with minimal hardware overhead relative to a dedicated PAM4 link. The SST transmitter achieves 1.2 Vpp output swing and employs LUT control of a 31-segment output DAC to implement 4/2-tap FFE in NRZ/PAM4 modes, respectively. Power efficiency is improved in the transmitter with an optimized quarter-rate serializer and a new low-overhead analog impedance control scheme is employed in the output stage to obviate additional impedance control segments. The presented DFE receiver utilizes a new single-clock phase two-stage regenerative comparator in the 2-bit flash ADCs to allow sufficient gain to support PAM4 DFE. Improved sensitivity is achieved in the direct feedback design with the multilevel first post-cursor ISI subtracted in the comparators and the remaining ISI cancelled in a preceding current integration summer. Overall, leveraging the proposed dual-mode SerDes architecture allows the support of multiple channel conditions and variable data rates with a single design solution.

APPENDIX

While pessimistic from a BER perspective, peak distortion analysis [37] provides a rapid approach to find the worst-case eye opening and is utilized to highlight the differences in ISI sensitivity between NRZ and PAM4 modulation at the same symbol rate. Fig. 23 shows a conceptual pulse response \( y(t) \) produced by sending an ideal pulse \( c(t) \) with duration \( T_b \) across a channel. This pulse response has the cursor value at \( t = t_0 \) and ISI terms at \( T_b \) offsets before and after this cursor instant.

First consider the NRZ modulation case, where there are two symbols, \( y_1 = y(t) \) and \( y_0 = -y(t) \). Assuming linearity, the worst case high and low levels, \( v_1 \) and \( v_0 \), respectively,
of the eye diagram at the sampling point, $t_o$, is calculated by

$$v_1 = y(t_o) - \sum_{i=-\infty}^{\infty} |y(t_o - iT_b)|$$

$$v_0 = -y(t_o) + \sum_{i=-\infty}^{\infty} |y(t_o - iT_b)|. \tag{1}$$

Thus, the NRZ PDA eye height, shown in Fig. 24(a), is

$$A_{NRZ} = v_1 - v_0 = 2 \left( y(t_o) - \sum_{i=-\infty}^{\infty} |y(t_o - iT_b)| \right). \tag{2}$$

Note that the $\sum_{i=-\infty}^{\infty} |y(t_o - iT_b)|$ term equals the sum of the absolute value of all post- and pre-cursor ISI values determined from the pulse response. This represents the maximum amount of ISI that can be added or subtracted from a symbol with the worst-case symbol sequence. In the common case where all ISI values are positive, a lone-pulse sequence of a single 1 preceded and followed by all 0s is the worst-case pattern that sets the minimum high level.

Now consider the PAM4 case, where there are four symbols, $y_{11} = y(t)$, $y_{10} = 1/3 \ y(t)$, $y_{01} = -1/3 \ y(t)$, and $y_{00} = -y(t)$. As shown in Fig. 24(b), assuming linearity this results in three eyes that are bounded by six levels which can be calculated by

$$v_{11} = y(t_o) - \sum_{i=-\infty}^{\infty} |y(t_o - iT_b)|$$

$$v_{10h} = \frac{1}{3} y(t_o) + \sum_{i=-\infty}^{\infty} |y(t_o - iT_b)|$$

$$v_{10l} = \frac{1}{3} y(t_o) - \sum_{i=-\infty}^{\infty} |y(t_o - iT_b)|$$

$$v_{01h} = -\frac{1}{3} y(t_o) + \sum_{i=-\infty}^{\infty} |y(t_o - iT_b)|$$

$$v_{01l} = -\frac{1}{3} y(t_o) - \sum_{i=-\infty}^{\infty} |y(t_o - iT_b)|$$

$$v_{00} = -y(t_o) + \sum_{i=-\infty}^{\infty} |y(t_o - iT_b)|. \tag{3}$$

Thus, the PAM4 PDA eye heights are

$$A_{PAM4} = v_{11} - v_{10h} = v_{10l} - v_{10h} = v_{10l} - v_{00} = 2 \left( \frac{1}{3} y(t_o) - \sum_{i=-\infty}^{\infty} |y(t_o - iT_b)| \right). \tag{4}$$

Note that although the ideal voltage margin with PAM4 modulation is 1/3 the ideal voltage margin with NRZ modulation, the PAM4 symbols suffer from the same amount of ISI $\sum_{i=-\infty}^{\infty} |y(t_o - iT_b)|$. While for the same data rate a PAM4 pulse response will often be much better than its NRZ counterpart for typical wireline channels, it is worth noting that this heightened PAM4 ISI sensitivity necessitates an increased level of ISI cancellation.

### References


Asghan Roshan-Zamir (S’14) received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Tehran, Tehran, Iran, in 2010 and 2013, respectively. He is currently pursuing the Ph.D. degree in electrical engineering with Texas A&M University, College Station, TX, USA. Since 2013, he has been a Research Assistant with the Analog and Mixed Signal Center, Texas A&M University. His current research interests include analog and mixed-signal integrated circuits, high-speed electrical and optical transceiver circuits, and high-speed clocking circuits.

Osama Elhadjidi (S’11–M’16) received the B.Sc. and M.Sc. degrees in electrical engineering from Ain Shams University, Cairo, Egypt, in 2004 and 2009, respectively, and the Ph.D. degree in electrical engineering from Texas A&M University, College Station, TX, USA, in 2015. From 2005 to 2010, he was a Development Engineer with Mentor Graphics, Cairo. In 2012, he joined Rambus, Chapel Hill, NC, USA, as a Design Intern. From 2013 to 2014, he was a Design Intern with Texas Instruments, Dallas, TX, USA. Since 2015, he has been with Qualcomm Technologies Inc., San Diego, CA, USA. His current research interests include high-speed analog, mixed-signal, and RF integrated circuit design.
Hae-Woong Yang (S’13) was born in Seoul, South Korea. He received the B.S. and M.E. degrees in electrical and computer engineering from Texas A&M University, College Station, TX, USA, in 2007 and 2009, respectively, where he is currently pursuing the Ph.D. degree with the Analog and Mixed Signal Center. His current research interests include low-power high-speed electrical link circuits, clock generation circuits, and signal integrity.

Mr. Yang was a co-recipient of the Student Best Paper Award in the 2014 Midwest Symposium on Circuits and Systems.

Samuel Palermo (S’98–M’07) received the B.S. and M.S. degrees in electrical engineering from Texas A&M University, College Station, TX, USA, in 1997 and 1999, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 2007. From 1999 to 2000, he was with Texas Instruments, Dallas, TX, USA, where he was involved in the design of mixed-signal integrated circuits for high-speed serial data communication. From 2006 to 2008, he was with Intel Corporation, Hillsboro, OR, USA, where he was involved in high-speed optical and electrical I/O architectures. In 2009, he joined the Electrical and Computer Engineering Department, Texas A&M University, where he is currently an Associate Professor. His current research interests include high-speed electrical and optical interconnect architectures, high-performance clocking circuits, and integrated sensor systems.

Dr. Palermo is a member of Eta Kappa Nu. He was a recipient of the 2013 NSF-CAREER Award, the Texas A&M University Department of Electrical and Computer Engineering Outstanding Professor Award in 2014, and the Engineering Faculty Fellow Award in 2015. He served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II from 2011 to 2015 and served on the IEEE CASS Board of Governors from 2011 to 2012. He is currently a Distinguished Lecturer of the IEEE Solid-State Circuits Society. He was a co-recipient of the Jack Raper Award for Outstanding Technology Directions Paper at the 2009 International Solid-State Circuits Conference, the Best Student Paper at the 2014 Midwest Symposium on Circuits and Systems, and the Best Student Paper at the 2016 Dallas Circuits and Systems Conference.